Supplementary Materials

A. Proof of Properties of Median (Section 5.1)

A.1. Proof of Lemma 1

For simplicity, denote $\theta_p := \theta_p(F)$ and $\hat{\theta}_p := \theta_p(\{X_i\}_{i=1}^m)$. Since F' is continuous and positive, for an ϵ, there exists a δ_1 such that $\mathbb{P}(X \leq \theta_p - \epsilon) = p - \delta_1$, where $\delta_1 \in (\epsilon l, \epsilon L)$. Then one has

$$\begin{aligned}\mathbb{P}\left(\hat{\theta}_p < \theta_p - \epsilon\right) &\overset{(a)}{=} \mathbb{P}\left(\sum_{i=1}^m \mathbb{1}_{\{X_i \leq \theta_p - \epsilon\}} \geq pm\right) = \mathbb{P}\left(\frac{1}{m}\sum_{i=1}^m \mathbb{1}_{\{X_i \leq \theta_p - \epsilon\}} \geq (p-\delta_1) + \delta_1\right) \\ &\overset{(b)}{\leq} \exp(-2m\delta_1^2) \leq \exp(-2m\epsilon^2 l^2),\end{aligned}$$

where (a) is due to the definition of the quantile function in (15) and (b) is due to the fact that $\mathbb{1}_{\{X_i \leq \theta_p - \epsilon\}} \sim \text{Bernoulli}(p - \delta_1)$ i.i.d., followed by the Hoeffding inequality. Similarly, one can show for some $\delta_2 \in (\epsilon l, \epsilon L)$,

$$\mathbb{P}\left(\hat{\theta}_p > \theta_p + \epsilon\right) \leq \exp(-2m\delta_2^2) \leq \exp(-2m\epsilon^2 l^2).$$

Combining these two inequalities, one has the conclusion.

A.2. Proof of Lemma 2

It suffices to show that

$$|X_{(k)} - Y_{(k)}| \leq \max_l |X_l - Y_l|, \quad \forall k = 1, \cdots, n. \tag{25}$$

Case 1: $k = n$, suppose $X_{(n)} = X_i$ and $Y_{(n)} = Y_j$, i.e., X_i is the largest among $\{X_l\}_{l=1}^n$ and Y_j is the largest among $\{Y_l\}_{l=1}^n$. Then we have either $X_j \leq X_i \leq Y_j$ or $Y_i \leq Y_j \leq X_i$. Hence,

$$|X_{(n)} - Y_{(n)}| = |X_i - Y_j| \leq \max\{|X_i - Y_i|, |X_j - Y_j|\}.$$

Case 2: $k = 1$, suppose that $X_{(1)} = X_i$ and $Y_{(1)} = Y_j$. Similarly

$$|X_{(1)} - Y_{(1)}| = |X_i - Y_j| \leq \max\{|X_i - Y_i|, |X_j - Y_j|\}.$$

Case 3: $1 < k < n$, suppose that $X_{(k)} = X_i$, $Y_{(k)} = Y_j$, and without loss of generality assume that $X_i < Y_j$ (if $X_i = Y_j$, $0 = |X_{(k)} - Y_{(k)}| \leq \max_l |X_l - Y_l|$ holds trivially). We show the conclusion by contradiction.

Assume $|X_{(k)} - Y_{(k)}| > \max_l |X_l - Y_l|$. Then one must have $Y_i < Y_j$ and $X_j > X_i$ and $i \neq j$. Moreover for any $p < k$ and $q > k$, the index of $X_{(p)}$ cannot be equal to the index of $Y_{(q)}$; otherwise the assumption is violated.

Thus, all $Y_{(q)}$ for $q > k$ must share the same index set with $X_{(p)}$ for $p > k$. However, X_j, which is larger than X_i (thus if $X_j = X_{(k')}$, then $k' > k$), shares the same index with Y_j, where $Y_j = Y_{(k)}$. This yields contradiction.

A.3. Proof of Lemma 3

Assume that sm is an integer. Since there are sm corrupted samples in total, one can select out at least $\lceil (\frac{1}{2} - s)m \rceil$ clean samples from the left half of ordered contaminated samples $\{\theta_{1/m}(\{X_i\}), \theta_{2/m}(\{X_i\}), \cdots, \theta_{1/2}(\{X_i\})\}$. Thus one has the left inequality. Furthermore, one can also select out at least $\lceil (\frac{1}{2} - s)m \rceil$ clean samples from the right half of ordered contaminated samples $\{\theta_{1/2}(\{X_i\}), \cdots, \theta_1(\{X_i\})\}$. One has the right inequality.

A.4. Proof of Lemma 4

First we introduce some general facts for the distribution of the product of two correlated standard Gaussian random variables (Donahue, 1964). Let $u \sim \mathcal{N}(0,1)$, $v \sim \mathcal{N}(0,1)$, and their correlation coefficient be $\rho \in [-1, 1]$. Then the density of uv is given by

$$\phi_\rho(x) = \frac{1}{\pi\sqrt{1-\rho^2}} \exp\left(\frac{\rho x}{1-\rho^2}\right) K_0\left(\frac{|x|}{1-\rho^2}\right), \quad x \neq 0,$$

where $K_0(\cdot)$ is the modified Bessel function of the second kind. Thus the density of $r = |uv|$ is

$$\psi_\rho(x) = \frac{1}{\pi\sqrt{1-\rho^2}} \left[\exp\left(\frac{\rho x}{1-\rho^2}\right) + \exp\left(-\frac{\rho x}{1-\rho^2}\right)\right] K_0\left(\frac{|x|}{1-\rho^2}\right), \quad x > 0, \tag{26}$$

for $|\rho| < 1$. If $|\rho| = 1$, r becomes a χ_1^2 random variable, with the density

$$\psi_{|\rho|=1}(x) = \frac{1}{\sqrt{2\pi}} x^{-1/2} \exp(-x/2), \quad x > 0.$$

It can be seen from (26) that the density of r only relates to the correlation coefficient $\rho \in [-1, 1]$.

Let $\theta_{1/2}(\psi_\rho)$ be the $1/2$ quantile (median) of the distribution $\psi_\rho(x)$, and $\psi_\rho(\theta_{1/2})$ be the value of the function ψ_ρ at the point $\theta_{1/2}(\psi_\rho)$. Although it is difficult to derive the analytical expressions of $\theta_{1/2}(\psi_\rho)$ and $\psi_\rho(\theta_{1/2})$ due to the complicated form of ψ_ρ in (26), due to the continuity of $\psi_\rho(x)$ and $\theta_{1/2}(\psi_\rho)$, we can calculate them numerically, as illustrated in Figure 4. From the numerical calculation, one can see that both $\psi_\rho(\theta_{1/2})$ and $\theta_{1/2}(\psi_\rho)$ are bounded from below and above

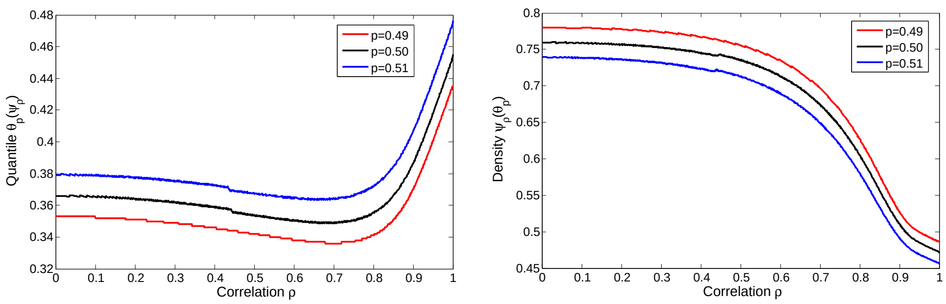

Figure 4. Quantiles and density at quantiles across ρ

for all $\rho \in [0, 1]$ ($\psi_\rho(\cdot)$ is symmetric over ρ, hence it is sufficient to consider $\rho \in [0, 1]$), satisfying

$$0.348 < \theta_{1/2}(\psi_\rho) < 0.455, \quad 0.47 < \psi_\rho(\theta_{1/2}) < 0.76. \tag{27}$$

B. Robust Initialization with Outliers (Section 5.2)

This section proves that the truncated spectral method provides a good initialization even if sm measurements are corrupted by arbitrary outliers as long as s is small.

Consider the model in (1). Lemma 3 yields

$$\theta_{\frac{1}{2}-s}(\{(\boldsymbol{a}_i^T\boldsymbol{x})^2\}) < \theta_{1/2}(\{y_i\}) < \theta_{\frac{1}{2}+s}(\{(\boldsymbol{a}_i^T\boldsymbol{x})^2\}). \tag{28}$$

Observe that $\boldsymbol{a}_i^T\boldsymbol{x} = \tilde{a}_{i1}^2\|\boldsymbol{x}\|^2$, where $\tilde{a}_{i1} = \boldsymbol{a}_i^T\boldsymbol{x}/\|\boldsymbol{x}\|$ is a standard Gaussian random variable. Thus $|\tilde{a}_{i1}|^2$ is a χ_1^2 random variable, whose cumulative distribution function is denoted as $K(x)$. Moreover by Lemma 1, for a small ϵ, one has $\left|\theta_{\frac{1}{2}-s}(\{|\tilde{a}_{i1}|^2\}) - \theta_{\frac{1}{2}-s}(K)\right| < \epsilon$ and $\left|\theta_{\frac{1}{2}+s}(\{|\tilde{a}_{i1}|^2\}) - \theta_{\frac{1}{2}+s}(K)\right| < \epsilon$ with probability $1 - 2\exp(-cm\epsilon^2)$ and c is a constant around 2×0.47^2 (c.f. Figure 4). We note that $\theta_{1/2}(K) = 0.455$ and both $\theta_{\frac{1}{2}-s}(K)$ and $\theta_{\frac{1}{2}+s}(K)$ can be arbitrarily close to $\theta_{\frac{1}{2}}(K)$ simultaneously as long as s is small enough (independent of n). Thus one has

$$\left(\theta_{\frac{1}{2}-s}(K) - \epsilon\right)\|\boldsymbol{x}\|^2 < \theta_{1/2}(\{y_i\}) < \left(\theta_{\frac{1}{2}+s}(K) + \epsilon\right)\|\boldsymbol{x}\|^2, \tag{29}$$

with probability at least $1 - \exp(-cm\epsilon^2)$. For the sake of simplicity, we introduce two new notations $\zeta_s := \theta_{\frac{1}{2}-s}(K)$ and $\zeta^s := \theta_{\frac{1}{2}+s}(K)$. Specifically for the instance of $s = 0.01$, one has $\zeta_s = 0.434$ and $\zeta^s = 0.477$. It is easy to see that $\zeta^s - \zeta_s$ can be arbitrarily small if s is small enough.

We first consider the case when $\|\boldsymbol{x}\| = 1$. On the event that (29) holds, the truncation function has the following bounds,

$$\mathbb{1}_{\{y_i \le \alpha_y^2\theta_{1/2}(\{y_i\})/0.455\}} \le \mathbb{1}_{\left\{y_i \le \alpha_y^2(\zeta^s+\epsilon)/0.455\right\}}$$
$$\mathbb{1}_{\{y_i \le \alpha_y^2\theta_{1/2}(\{y_i\})/0.455\}} \ge \mathbb{1}_{\left\{y_i \le \alpha_y^2(\zeta_s-\epsilon)/0.455\right\}}.$$

On the other hand, denote the support of the outliers as S, we have

$$\boldsymbol{Y} = \frac{1}{m}\sum_{i\notin S} \boldsymbol{a}_i\boldsymbol{a}_i^T(\boldsymbol{a}_i^T\boldsymbol{x})^2 \mathbb{1}_{\{(\boldsymbol{a}_i^T\boldsymbol{x})^2 \le \alpha_y^2\theta_{1/2}(\{y_i\})/0.455\}} + \frac{1}{m}\sum_{i\in S} \boldsymbol{a}_i\boldsymbol{a}_i^T y_i \mathbb{1}_{\{y_i \le \alpha_y^2\theta_{1/2}(\{y_i\})/0.455\}}.$$

Consequently, one can bound $\boldsymbol{Y}$ as

$$\begin{aligned}\boldsymbol{Y}_1 :=& \frac{1}{m}\sum_{i\notin S} \boldsymbol{a}_i\boldsymbol{a}_i^T(\boldsymbol{a}_i^T\boldsymbol{x})^2 \mathbb{1}_{\{(\boldsymbol{a}_i^T\boldsymbol{x})^2 \le \alpha_y^2(\zeta_s-\epsilon)/0.455\}} \prec \boldsymbol{Y}\\ \prec& \frac{1}{m}\sum_{i\notin S} \boldsymbol{a}_i\boldsymbol{a}_i^T(\boldsymbol{a}_i^T\boldsymbol{x})^2 \mathbb{1}_{\{(\boldsymbol{a}_i^T\boldsymbol{x})^2 \le \alpha_y^2(\zeta^s+\epsilon)/0.455\}} + \frac{1}{m}\sum_{i\in S} \boldsymbol{a}_i\boldsymbol{a}_i^T \alpha_y^2(\zeta^s+\epsilon)/0.455 =: \boldsymbol{Y}_2,\end{aligned}$$

where we have

$$\mathbb{E}[\boldsymbol{Y}_1] = (1-s)(\beta_1\boldsymbol{x}\boldsymbol{x}^T + \beta_2\boldsymbol{I}), \quad \mathbb{E}[\boldsymbol{Y}_2] = (1-s)(\beta_3\boldsymbol{x}\boldsymbol{x}^T + \beta_4\boldsymbol{I}) + s\alpha_y^2\frac{(\zeta^s+\epsilon)}{0.455}\boldsymbol{I}, \tag{30}$$

with $\beta_1 := \mathbb{E}[\xi^4\mathbb{1}_{\left\{|\xi|\le\alpha_y\sqrt{(\zeta_s-\epsilon)/0.455}\right\}}] - \mathbb{E}[\xi^2\mathbb{1}_{\left\{|\xi|\le\alpha_y\sqrt{(\zeta_s-\epsilon)/0.455}\right\}}]$, $\beta_2 := \mathbb{E}[\xi^2\mathbb{1}_{\left\{|\xi|\le\alpha_y\sqrt{(\zeta_s-\epsilon)/0.455}\right\}}]$ and $\beta_3 := \mathbb{E}[\xi^4\mathbb{1}_{\left\{|\xi|\le\alpha_y\sqrt{(\zeta^s+\epsilon)/0.455}\right\}}] - \mathbb{E}[\xi^2\mathbb{1}_{\left\{|\xi|\le\alpha_y\sqrt{(\zeta^s+\epsilon)/0.455}\right\}}]$, $\beta_4 := \mathbb{E}[\xi^2\mathbb{1}_{\left\{|\xi|\le\alpha_y\sqrt{(\zeta^s+\epsilon)/0.455}\right\}}]$, assuming $\xi \sim \mathcal{N}(0,1)$.

Applying standard results on random matrices with non-isotropic sub-Gaussian rows (Vershynin, 2012, equation (5.26)) and noticing that $\boldsymbol{a}_i\boldsymbol{a}_i^T(\boldsymbol{a}_i^T\boldsymbol{x})^2\mathbb{1}_{\{|\boldsymbol{a}_i^T\boldsymbol{x}|\le c\}}$ can be rewritten as $\boldsymbol{b}_i\boldsymbol{b}_i^T$ for some sub-Gaussian vector $\boldsymbol{b}_i := \boldsymbol{a}_i(\boldsymbol{a}_i^T\boldsymbol{x})\mathbb{1}_{\{|\boldsymbol{a}_i^T\boldsymbol{x}|\le c\}}$, one can deduce

$$\|\boldsymbol{Y}_1 - \mathbb{E}[\boldsymbol{Y}_1]\| \le \delta, \quad \|\boldsymbol{Y}_2 - \mathbb{E}[\boldsymbol{Y}_2]\| \le \delta \tag{31}$$

with probability $1-\exp(-\Omega(m))$, provided that m/n exceeds some large constant. Besides, when ϵ and s are sufficiently small, one further has $\|\mathbb{E}[\boldsymbol{Y}_1] - \mathbb{E}[\boldsymbol{Y}_2]\| \le \delta$. Putting these together, one has

$$\|\boldsymbol{Y} - (1-s)(\beta_1\boldsymbol{x}\boldsymbol{x}^T + \beta_2\boldsymbol{I})\| \le 3\delta. \tag{32}$$

Let $\tilde{\boldsymbol{z}}^{(0)}$ be the normalized leading eigenvector of $\boldsymbol{Y}$. Repeating the same argument as in (Candès et al., 2015, Section 7.8) and taking δ, ϵ to be sufficiently small, one has

$$\text{dist}(\tilde{\boldsymbol{z}}^{(0)}, \boldsymbol{x}) \le \tilde{\delta}, \tag{33}$$

for a given $\tilde{\delta} > 0$, as long as m/n exceeds some large constant.

Furthermore let $\boldsymbol{z}^{(0)} = \sqrt{\mathsf{med}\{y_i\}/0.455}\,\tilde{\boldsymbol{z}}^{(0)}$ to handle cases $\|\boldsymbol{x}\| \ne 1$. By the bound (29), one has

$$\left|\frac{\mathsf{med}(\{y_i\})}{0.455} - \|\boldsymbol{x}\|^2\right| \le \max\left\{\left|\frac{\zeta_s-\epsilon}{0.455} - 1\right|, \left|\frac{\zeta^s+\epsilon}{0.455} - 1\right|\right\}\|\boldsymbol{x}\|^2 \le \frac{\zeta^s - \zeta_s + \epsilon}{0.455}\|\boldsymbol{x}\|^2 \tag{34}$$

Thus

$$\text{dist}(\boldsymbol{z}^{(0)}, \boldsymbol{x}) \le \frac{\zeta^s - \zeta_s + \epsilon}{0.455}\|\boldsymbol{x}\| + \tilde{\delta}\|\boldsymbol{x}\| \le \frac{1}{11}\|\boldsymbol{x}\| \tag{35}$$

as long as s is a small enough constant.

C. Geometric Convergence for Noise-free Model (Proof of Corollary 2)

After obtaining a good initialization, the central idea to establish geometric convergence is to show that the truncated gradient $\nabla\ell_{tr}(\boldsymbol{z})$ in the neighborhood of the global optima satisfies the *Regularity Condition* $RC(\mu, \lambda, \epsilon)$ defined in Definition 2. We show this by two steps. Step 1 establishes a key concentration property for the sample median used in the truncation rule, which is then subsequently exploited to prove RC in Step 2.

C.1. Proof of Concentration Property for Sample Median

We show that the sample median used in the truncation rule concentrates at the level $\|\boldsymbol{z}-\boldsymbol{x}\|\|\boldsymbol{z}\|$ as stated in the following proposition. Along the way, we also establish that the sample quantiles around the median are also concentrated at the level $\|\boldsymbol{z}-\boldsymbol{x}\|\|\boldsymbol{z}\|$.

Proposition 2 (Refined version of Proposition 1)**.** *Fix $\epsilon \in (0,1)$. If $m > c_0(\epsilon^{-2}\log\frac{1}{\epsilon})n\log n$, then with probability at least $1-c_1\exp(-c_2 m\epsilon^2)$,*

$$(0.65-\epsilon)\|\boldsymbol{z}\|\|\boldsymbol{h}\| \le \mathsf{med}\{|(\boldsymbol{a}_i^T\boldsymbol{x})^2-(\boldsymbol{a}_i^T\boldsymbol{z})^2|\} \le (0.91+\epsilon)\|\boldsymbol{z}\|\|\boldsymbol{h}\|, \tag{36}$$

$$(0.63-\epsilon)\|\boldsymbol{z}\|\|\boldsymbol{h}\| \le \theta_{0.49},\theta_{0.51}\{|(\boldsymbol{a}_i^T\boldsymbol{x})^2-(\boldsymbol{a}_i^T\boldsymbol{z})^2|\} \le (0.95+\epsilon)\|\boldsymbol{z}\|\|\boldsymbol{h}\|, \tag{37}$$

hold for all $\boldsymbol{x},\boldsymbol{z}$ with $\|\boldsymbol{x}-\boldsymbol{z}\| < 1/11\|\boldsymbol{z}\|$, where $\boldsymbol{h}:=\boldsymbol{z}-\boldsymbol{x}$.

Proof. We first show for a fixed pair $\boldsymbol{z}$ and $\boldsymbol{x}$, (36) and (37) hold with high probability.

Let $r_i = |(\boldsymbol{a}_i^T\boldsymbol{x})^2-(\boldsymbol{a}_i^T\boldsymbol{z})^2|$. Then r_i's are i.i.d. copies of a random variable r, where $r=|(\boldsymbol{a}^T\boldsymbol{x})^2-(\boldsymbol{a}^T\boldsymbol{z})^2|$ with the entries of $\boldsymbol{a}$ composed of i.i.d. standard Gaussian random variables. Note that the distribution of r is fixed once given $\boldsymbol{h}$ and $\boldsymbol{z}$.

Let $\boldsymbol{x}(1)$ denote the first element of a generic vector $\boldsymbol{x}$, and $\boldsymbol{x}_{-1}$ denote the remaining vector of $\boldsymbol{x}$ after eliminating the first element. Let $\boldsymbol{U}_h$ be an orthonormal matrix with first row being $\boldsymbol{h}^T/\|\boldsymbol{h}\|$, and $\tilde{\boldsymbol{a}}=\boldsymbol{U}_h\boldsymbol{a}$, $\tilde{\boldsymbol{z}}=\boldsymbol{U}_h\boldsymbol{z}$. Similarly define $\boldsymbol{U}_{\tilde{z}_{-1}}$ and let $\tilde{\boldsymbol{b}}=\boldsymbol{U}_{\tilde{z}_{-1}}\tilde{\boldsymbol{a}}_{-1}$. Then $\tilde{\boldsymbol{a}}(1)$ and $\tilde{\boldsymbol{b}}(1)$ are independent standard normal random variables. We further express r as follows.

$$\begin{aligned}
r &= |(\boldsymbol{a}^T\boldsymbol{z})^2-(\boldsymbol{a}^T\boldsymbol{x})^2| \\
&= |(2\boldsymbol{a}^T\boldsymbol{z}-\boldsymbol{a}^T\boldsymbol{h})(\boldsymbol{a}^T\boldsymbol{h})| \\
&= |(2\tilde{\boldsymbol{a}}^T\tilde{\boldsymbol{z}}-\tilde{\boldsymbol{a}}(1)\|\boldsymbol{h}\|)(\tilde{\boldsymbol{a}}(1)\|\boldsymbol{h}\|)| \\
&= |(2\boldsymbol{h}^T\boldsymbol{z}-\|\boldsymbol{h}\|^2)\tilde{\boldsymbol{a}}(1)^2+2(\tilde{\boldsymbol{a}}_{-1}^T\tilde{\boldsymbol{z}}_{-1})(\tilde{\boldsymbol{a}}(1)\|\boldsymbol{h}\|)| \\
&= |(2\boldsymbol{h}^T\boldsymbol{z}-\|\boldsymbol{h}\|^2)\tilde{\boldsymbol{a}}(1)^2+2\tilde{\boldsymbol{b}}(1)\|\tilde{\boldsymbol{z}}_{-1}\|\tilde{\boldsymbol{a}}(1)\|\boldsymbol{h}\|| \\
&= |(2\boldsymbol{h}^T\boldsymbol{z}-\|\boldsymbol{h}\|^2)\tilde{\boldsymbol{a}}(1)^2+2\sqrt{\|\boldsymbol{z}\|^2-\tilde{\boldsymbol{z}}(1)^2}\,\tilde{\boldsymbol{a}}(1)\tilde{\boldsymbol{b}}(1)\|\boldsymbol{h}\|| \\
&= \left|\left(2\frac{\boldsymbol{h}^T\boldsymbol{z}}{\|\boldsymbol{h}\|\|\boldsymbol{z}\|}-\frac{\|\boldsymbol{h}\|}{\|\boldsymbol{z}\|}\right)\tilde{\boldsymbol{a}}(1)^2+2\sqrt{1-\left(\frac{\boldsymbol{h}^T\boldsymbol{z}}{\|\boldsymbol{h}\|\|\boldsymbol{z}\|}\right)^2}\,\tilde{\boldsymbol{a}}(1)\tilde{\boldsymbol{b}}(1)\right|\cdot\|\boldsymbol{h}\|\|\boldsymbol{z}\| \\
&=: \left|(2\cos(\omega)-t)\tilde{\boldsymbol{a}}(1)^2+2\sqrt{1-\cos^2(\omega)}\,\tilde{\boldsymbol{a}}(1)\tilde{\boldsymbol{b}}(1)\right|\cdot\|\boldsymbol{h}\|\|\boldsymbol{z}\| \\
&=: |u\tilde{v}|\cdot\|\boldsymbol{h}\|\|\boldsymbol{z}\|
\end{aligned}$$

where ω is the angle between $\boldsymbol{h}$ and $\boldsymbol{z}$, and $t=\|\boldsymbol{h}\|/\|\boldsymbol{z}\| < 1/11$. Consequently, $u=\tilde{\boldsymbol{a}}(1)\sim\mathcal{N}(0,1)$ and $\tilde{v}=(2\cos(\omega)-t)\tilde{\boldsymbol{a}}(1)+2|\sin(\omega)|\tilde{\boldsymbol{b}}(1)$ is also a Gaussian random variable with variance $3.6<\mathsf{Var}(\tilde{v})<4$ under the assumption $t<1/11$.

Let $v=\tilde{v}/\sqrt{\mathsf{Var}(\tilde{v})}$, then $v\sim\mathcal{N}(0,1)$. Furthermore, let $r'=|uv|$. Denote the density function of r' as $\psi_\rho(\cdot)$ and the $1/2$-quantile point of r' as $\theta_{1/2}(\psi_\rho)$. By Lemma 4, we have

$$0.47<\psi_\rho(\theta_{1/2})<0.76, \quad 0.348<\theta_{1/2}(\psi_\rho)<0.455. \tag{38}$$

By Lemma 1, we have with probability at least $1-2\exp(-cm\epsilon^2)$ (here c is around 2×0.47^2),

$$0.348-\epsilon<\mathsf{med}(\{r_i'\}_{i=1}^m)<0.455+\epsilon. \tag{39}$$

The same arguments carry over to other quantiles $\theta_{0.49}(\{r_i'\})$ and $\theta_{0.51}(\{r_i'\})$. From Figure. 4, we observe that for $\rho\in[0,1]$

$$0.45<\psi_\rho(\theta_{0.49}),\psi_\rho(\theta_{0.51})<0.78, \quad 0.336<\theta_{0.49}(\psi_\rho),\theta_{0.51}(\psi_\rho)<0.477 \tag{40}$$

and then we have with probability at least $1-2\exp(-cm\epsilon^2)$ (here c is around 2×0.45^2),

$$0.336-\epsilon < \theta_{0.49}(\{r'_m\}), \theta_{0.51}(\{r'_m\}) < 0.477+\epsilon. \tag{41}$$

Hence, by multiplying back $\sqrt{\mathsf{Var}(\tilde{v})}$, we have with probability $1-2\exp(-cm\epsilon^2)$,

$$(0.65-\epsilon)\|\boldsymbol{z}-\boldsymbol{x}\|\|\boldsymbol{z}\| \le \mathsf{med}\left(\left\{|(\boldsymbol{a}_i^T\boldsymbol{z})^2-(\boldsymbol{a}_i^T\boldsymbol{x})^2|\right\}\right) \le (0.91+\epsilon)\|\boldsymbol{z}-\boldsymbol{x}\|\|\boldsymbol{z}\|, \tag{42}$$

$$(0.63-\epsilon)\|\boldsymbol{z}-\boldsymbol{x}\|\|\boldsymbol{z}\| \le \theta_{0.49}, \theta_{0.51}\left(\left\{|(\boldsymbol{a}_i^T\boldsymbol{z})^2-(\boldsymbol{a}_i^T\boldsymbol{x})^2|\right\}\right) \le (0.95+\epsilon)\|\boldsymbol{z}-\boldsymbol{x}\|\|\boldsymbol{z}\|. \tag{43}$$

We note that, to keep notation simple, c and ϵ may vary line by line within constant factors.

Up to now, we proved for any fixed $\boldsymbol{z}$ and $\boldsymbol{x}$, the median or neighboring quantiles of $\left\{|(\boldsymbol{a}_i^T\boldsymbol{z})^2-(\boldsymbol{a}_i^T\boldsymbol{x})^2|\right\}$ are upper and lower bounded by $\|\boldsymbol{z}-\boldsymbol{x}\|\|\boldsymbol{z}\|$ times constant factors. To prove (36) and (37) for all $\boldsymbol{z}$ and $\boldsymbol{x}$ with $\|\boldsymbol{z}-\boldsymbol{x}\| \le \frac{1}{11}\|\boldsymbol{z}\|$, we use the net covering argument. Still we argue for median first and the same arguments carry over to other quantiles smoothly.

To proceed, we restate (42) as

$$(0.65-\epsilon) \le \mathsf{med}\left(\left\{\left|\left(\frac{2(\boldsymbol{a}_i^T\boldsymbol{z})}{\|\boldsymbol{z}\|}-\frac{\boldsymbol{a}_i^T\boldsymbol{h}}{\|\boldsymbol{h}\|}\frac{\|\boldsymbol{h}\|}{\|\boldsymbol{z}\|}\right)\frac{\boldsymbol{a}_i^T\boldsymbol{h}}{\|\boldsymbol{h}\|}\right|\right\}\right) \le (0.91+\epsilon), \tag{44}$$

holds with probability at least $1-2\exp(-cm\epsilon^2)$ for a given pair $\boldsymbol{h},\boldsymbol{z}$ satisfying $\|\boldsymbol{h}\|/\|\boldsymbol{z}\| \le 1/11$.

Let $\tau=\epsilon/(6n+6m)$, and let $\mathcal{S}_\tau$ be a τ-net covering the unit sphere, $\mathcal{L}_\tau$ be a τ-net covering a line with length $1/11$, and set

$$\mathcal{N}_\tau = \{(\boldsymbol{z}_0,\boldsymbol{h}_0,t_0) : (\boldsymbol{z}_0,\boldsymbol{h}_0,t_0) \in \mathcal{S}_\tau\times\mathcal{S}_\tau\times\mathcal{L}_\tau\}. \tag{45}$$

One has cardinality bound (i.e., the upper bound on the covering number) $|\mathcal{N}_\tau| \le (1+2/\tau)^{2n}/(11\tau) < (1+2/\tau)^{2n+1}$. Taking the union bound we have

$$(0.65-\epsilon) \le \mathsf{med}\left(\left\{|2(\boldsymbol{a}_i^T\boldsymbol{z}_0)-(\boldsymbol{a}_i^T\boldsymbol{h}_0)t_0||\boldsymbol{a}_i^T\boldsymbol{h}_0|\right\}\right) \le (0.91+\epsilon), \quad \forall(\boldsymbol{z}_0,\boldsymbol{h}_0,t_0)\in\mathcal{N}_\epsilon \tag{46}$$

with probability at least $1-(1+2/\tau)^{2n+1}\exp(-cm\epsilon^2)$.

We next argue that (46) holds with probability $1-c_1\exp(-c_2m\epsilon^2)$ for some constants c_1, c_2 as long as $m \ge c_0(\epsilon^{-2}\log\epsilon^{-1})n\log n$ for sufficient large constant c_0. To prove this claim, we first observe

$$(1+2/\tau)^{2n+1} \asymp \exp(2n(\log(n+m)+\log 12+\log(1/\epsilon))) \asymp \exp(2n(\log m)).$$

We note that once ϵ is chosen, it is fixed in the whole proof and does not scale with m or n. For simplicity, assume that $\epsilon < 1/e$. Fix some positive constant $c' < c - c_2$. It then suffices to show that there exist large constant c_0 such that if $m \ge c_0(\epsilon^{-2}\log\epsilon^{-1})n\log n$, then

$$2n\log m < c'm\epsilon^2. \tag{47}$$

For any fixed n, if (47) holds for some m and $m > (2/c')\epsilon^{-2}n$, then (47) always holds for larger m, because

$$\begin{aligned}2n\log(m+1) &= 2n\log m + 2n(\log(m+1)-\log m) = 2n\log m + \frac{2n}{m}\log(1+\frac{1}{m})^m \\ &\le 2n\log m + \frac{2n}{m} \le c'm\epsilon^2 + c'\epsilon^2 = c'(m+1)\epsilon^2.\end{aligned}$$

Next, for any n, we can always find a c_0 such that (47) holds for $m=c_0(\epsilon^{-2}\log\epsilon^{-1})n\log n$. Such c_0 can be easily found for large n, i.e., $c_0=4/c'$ is a valid option if

$$(4/c')(\epsilon^{-2}\log\epsilon^{-1})n\log n < n^2. \tag{48}$$

Moreover, since the number of n that violates (48) is finite, the maximum over all such c_0 serves the purpose.

Next, one needs to bound

$$\left|\mathsf{med}\left(\left\{|2(\boldsymbol{a}_i^T\boldsymbol{z}_0)-(\boldsymbol{a}_i^T\boldsymbol{h}_0)t_0||\boldsymbol{a}_i^T\boldsymbol{h}_0|\right\}\right)-\mathsf{med}\left(\left\{|2(\boldsymbol{a}_i^T\boldsymbol{z})-(\boldsymbol{a}_i^T\boldsymbol{h})t||\boldsymbol{a}_i^T\boldsymbol{h}|\right\}\right)\right|$$

for any $\|\boldsymbol{z}-\boldsymbol{z}_0\|<\tau$, $\|\boldsymbol{z}-\boldsymbol{z}_0\|<\tau$ and $\|t-t_0\|<\tau$.

By Lemma 2 and the relation $\big||x|-|y|\big|\leq|x-y|$, we have

$$\begin{aligned}
&\left|\mathsf{med}\left(\left\{|2(\boldsymbol{a}_i^T\boldsymbol{z}_0)-(\boldsymbol{a}_i^T\boldsymbol{h}_0)t_0||\boldsymbol{a}_i^T\boldsymbol{h}_0|\right\}\right)-\mathsf{med}\left(\left\{|2(\boldsymbol{a}_i^T\boldsymbol{z})-(\boldsymbol{a}_i^T\boldsymbol{h})t||\boldsymbol{a}_i^T\boldsymbol{h}|\right\}\right)\right|\\
&\leq\max_{i\in[m]}\left|\left(2(\boldsymbol{a}_i^T\boldsymbol{z}_0)-(\boldsymbol{a}_i^T\boldsymbol{h}_0)t_0\right)(\boldsymbol{a}_i^T\boldsymbol{h}_0)-\left(2(\boldsymbol{a}_i^T\boldsymbol{z})-(\boldsymbol{a}_i^T\boldsymbol{h})t\right)(\boldsymbol{a}_i^T\boldsymbol{h})\right|\\
&\leq\max_{i\in[m]}\left|\left(2(\boldsymbol{a}_i^T\boldsymbol{z}_0)-(\boldsymbol{a}_i^T\boldsymbol{h}_0)t_0\right)(\boldsymbol{a}_i^T\boldsymbol{h}_0)-\left(2(\boldsymbol{a}_i^T\boldsymbol{z})-(\boldsymbol{a}_i^T\boldsymbol{h})t\right)(\boldsymbol{a}_i^T\boldsymbol{h}_0)\right|\\
&\quad+\max_{i\in[m]}\left|\left(2(\boldsymbol{a}_i^T\boldsymbol{z})-(\boldsymbol{a}_i^T\boldsymbol{h})t\right)(\boldsymbol{a}_i^T\boldsymbol{h}_0)-\left(2(\boldsymbol{a}_i^T\boldsymbol{z})-(\boldsymbol{a}_i^T\boldsymbol{h})t\right)(\boldsymbol{a}_i^T\boldsymbol{h})\right|\\
&\leq\max_{i\in[m]}\left(\left|2\boldsymbol{a}_i^T(\boldsymbol{z}_0-\boldsymbol{z})\right|+\left|(\boldsymbol{a}_i^T\boldsymbol{h}_0)t_0-(\boldsymbol{a}_i^T\boldsymbol{h})t\right|\right)\left|\boldsymbol{a}_i^T\boldsymbol{h}_0\right|+\max_{i\in[m]}\left|2(\boldsymbol{a}_i^T\boldsymbol{z})-(\boldsymbol{a}_i^T\boldsymbol{h})t\right|\left|\boldsymbol{a}_i^T(\boldsymbol{h}_0-\boldsymbol{h})\right|\\
&\leq\max_{i\in[m]}\|\boldsymbol{a}_i\|^2(3+t)\tau+\max_{i\in[m]}\|\boldsymbol{a}_i\|^2(2+t)\tau\\
&\leq\max_{i\in[m]}\|\boldsymbol{a}_i\|^2(5+2t)\tau
\end{aligned}$$

On the event $E_1:=\left\{\max_{i\in[m]}\|\boldsymbol{a}_i\|^2\leq m+n\right\}$, one can show that

$$\left|\mathsf{med}\left(\left\{|2(\boldsymbol{a}_i^T\boldsymbol{z}_0)-(\boldsymbol{a}_i^T\boldsymbol{h}_0)t_0||\boldsymbol{a}_i^T\boldsymbol{h}_0|\right\}\right)-\mathsf{med}\left(\left\{|2(\boldsymbol{a}_i^T\boldsymbol{z})-(\boldsymbol{a}_i^T\boldsymbol{h})t||\boldsymbol{a}_i^T\boldsymbol{h}|\right\}\right)\right|<6(m+n)\tau<\epsilon. \tag{49}$$

We claim that E_1 holds with probability at least $1-m\exp(-m/8)$ if $m>n$. This can be argued as follows. Notice that $\|\boldsymbol{a}_i\|^2=\sum_{j=1}^n\boldsymbol{a}_i(j)^2$, where $\boldsymbol{a}_i(j)$ is the j th element of $\boldsymbol{a}_i$. In other words, $\|\boldsymbol{a}_i\|^2$ is a sum of n i.i.d. χ_1^2 random variables. Applying the Bernstein-type inequality (Corollary5.17 Vershynin) and observing that the sub-exponential norm of χ_1^2 is smaller than 2, we have

$$\mathbb{P}\left\{\|\boldsymbol{a}_i\|^2\geq m+n\right\}\leq\exp(-m/8). \tag{50}$$

Then a union bound concludes the claim.

Note that (46) holds on an event E_2, which has probability $1-c_1\exp(-c_2m\epsilon^2)$ as long as $m\geq c_0(\epsilon^{-2}\log\frac{1}{\epsilon})n\log n$. On the intersection of E_1 and E_2, (36) holds.

The net covering arguments can also carry over to show that (37) holds for all $\boldsymbol{x}$ and $\boldsymbol{z}$ obeying $\|\boldsymbol{x}-\boldsymbol{z}\|\leq\frac{1}{11}\|\boldsymbol{z}\|$. ☐

C.2. Proof of RC

Following Proposition 2, we choose some small ϵ (i.e. $\epsilon<0.03$), then with probability at least $1-\exp(-\Omega(m))$,

$$0.6\|\boldsymbol{z}-\boldsymbol{x}\|\|\boldsymbol{z}\|\leq\mathsf{med}(\{|(\boldsymbol{a}_i^T\boldsymbol{x})^2-(\boldsymbol{a}_i^T\boldsymbol{z})^2|\})\leq1.0\|\boldsymbol{z}-\boldsymbol{x}\|\|\boldsymbol{z}\| \tag{51}$$

holds for all $\boldsymbol{z}$ and $\boldsymbol{x}$ satisfying $\|\boldsymbol{h}\|\leq1/11\|\boldsymbol{z}\|$. For each i, we introduce two new events

$$\mathcal{E}_3^i:=\{\left|(\boldsymbol{a}_i^T\boldsymbol{x})^2-(\boldsymbol{a}_i^T\boldsymbol{z})^2\right|\leq0.6\alpha_h\|\boldsymbol{h}\|\cdot|\boldsymbol{a}_i^T\boldsymbol{z}|\}, \tag{52}$$

$$\mathcal{E}_4^i:=\{\left|(\boldsymbol{a}_i^T\boldsymbol{x})^2-(\boldsymbol{a}_i^T\boldsymbol{z})^2\right|\leq1.0\alpha_h\|\boldsymbol{h}\|\cdot|\boldsymbol{a}_i^T\boldsymbol{z}|\}. \tag{53}$$

Conditioned on (51), the following inclusion property

$$\mathcal{E}_3^i\subseteq\mathcal{E}_2^i\subseteq\mathcal{E}_4^i \tag{54}$$

holds for all i, where $\mathcal{E}_2^i$ is defined in Algorithm 1. It is easier to work with these new events because $\mathcal{E}_3^i$'s (resp. $\mathcal{E}_4^i$'s) are statistically independent for any fixed $\boldsymbol{x}$ and $\boldsymbol{z}$. To further decouple the quadratic inequalities in $\mathcal{E}_3^i$ and $\mathcal{E}_4^i$ into linear inequalities, we introduce two more events and states their properties in the following lemma.

Lemma 5 (Lemma 3 in (Chen & Candès, 2015))**.** *For any $\gamma > 0$, define*

$$\mathcal{D}_\gamma^i := \{\left|(\boldsymbol{a}_i^* \boldsymbol{x})^2 - (\boldsymbol{a}_i^* \boldsymbol{z})^2\right| \le \gamma \|\boldsymbol{h}\| |\boldsymbol{a}_i^* \boldsymbol{z}|\}, \tag{55}$$

$$\mathcal{D}_\gamma^{i,1} := \left\{ \frac{|\boldsymbol{a}_i^* \boldsymbol{h}|}{\|\boldsymbol{h}\|} \le \gamma \right\}, \tag{56}$$

$$\mathcal{D}_\gamma^{i,2} := \left\{ \left| \frac{\boldsymbol{a}_i^* \boldsymbol{h}}{\|\boldsymbol{h}\|} - \frac{2\boldsymbol{a}_i^* \boldsymbol{z}}{\|\boldsymbol{h}\|} \right| \le \gamma \right\}. \tag{57}$$

On the event $\mathcal{E}_1^i$ defined in Algorithm 1, the quadratic inequality specifying $\mathcal{D}_\gamma^i$ implicates that $\boldsymbol{a}_i^T \boldsymbol{h}$ belongs to two intervals centered around 0 and $2\boldsymbol{a}_i^T \boldsymbol{z}$, respectively, i.e. $\mathcal{D}_\gamma^{i,1}$ and $\mathcal{D}_\gamma^{i,2}$. The following inclusion property holds

$$\left(\mathcal{D}^{i,1}_{\frac{\gamma}{1+\sqrt{2}}} \cap \mathcal{E}_1^i\right) \cup \left(\mathcal{D}^{i,2}_{\frac{\gamma}{1+\sqrt{2}}} \cap \mathcal{E}_1^i\right) \subseteq \mathcal{D}_\gamma^i \cap \mathcal{E}_1^i \subseteq \left(\mathcal{D}_\gamma^{i,1} \cap \mathcal{E}_1^i\right) \cup \left(\mathcal{D}_\gamma^{i,2} \cap \mathcal{E}_1^i\right). \tag{58}$$

Using Lemma 2, we can establish that $-\left\langle \frac{1}{m}\nabla \ell_{tr}(\boldsymbol{z}), \boldsymbol{h} \right\rangle$ is lower bounded on the order of $\|\boldsymbol{h}\|^2$, as in Proposition 3, and that $\left\| \frac{1}{m}\nabla \ell_{tr}(\boldsymbol{z}) \right\|$ is upper bounded on the order of $\|\boldsymbol{h}\|$, as in Proposition 4.

Proposition 3 (Adapted version of Proposition 2 of (Chen & Candès, 2015))**.** *Consider the noise-free measurements $y_i = |\boldsymbol{a}_i^T \boldsymbol{x}|^2$ and any fixed constant $\epsilon > 0$. Under the condition (10), if $m > c_0 n \log n$, then with probability at least $1 - c_1 \exp(-c_2 m)$,*

$$-\left\langle \frac{1}{m} \nabla \ell_{tr}(\boldsymbol{z}), \boldsymbol{h} \right\rangle \ge 2\left\{1.99 - 2(\zeta_1 + \zeta_2) - \sqrt{8/\pi}\alpha_h^{-1} - \epsilon\right\} \|\boldsymbol{h}\|^2 \tag{59}$$

holds uniformly over all $\boldsymbol{x}, \boldsymbol{z} \in \mathbb{R}^n$ satisfying

$$\frac{\|\boldsymbol{h}\|}{\|\boldsymbol{z}\|} \le \min\left\{ \frac{1}{11}, \frac{\alpha_l}{\alpha_h}, \frac{\alpha_l}{6}, \frac{\sqrt{98/3}(\alpha_l)^2}{2\alpha_u + \alpha_l} \right\}, \tag{60}$$

where $c_0, c_1, c_2 > 0$ are some universal constants, and ζ_1, ζ_2 are defined in (10).

The proof of Proposition 3 adapts the proof of Proposition 2 of (Chen & Candès, 2015), by properly setting parameters based on the properties of sample median. For completeness, we include a short outline of the proof in Appendix F.

Proposition 4 (Lemma 7 of (Chen & Candès, 2015))**.** *Under the same condition as in Proposition 3, if $m > c_0 n \log n$, then there exist some constants, $c_1, c_2 > 0$ such that with probability at least $1 - c_1 \exp(-c_2 m)$,*

$$\left\| \frac{1}{m} \nabla \ell_{tr}(\boldsymbol{z}) \right\| \le (1 + \delta) \cdot 4\sqrt{1.02 + 2/\alpha_h} \|\boldsymbol{h}\| \tag{61}$$

holds uniformly over all $\boldsymbol{x}, \boldsymbol{z} \in \mathbb{R}^n$ satisfying

$$\frac{\|\boldsymbol{h}\|}{\|\boldsymbol{z}\|} \le \min\left\{ \frac{1}{11}, \frac{\alpha_l}{\alpha_h}, \frac{\alpha_l}{6}, \frac{\sqrt{98/3}(\alpha_l)^2}{2\alpha_u + \alpha_l} \right\}, \tag{62}$$

where δ can be arbitrarily small as long as m/n sufficiently large.

With these two propositions, RC is guaranteed by setting $\mu < \mu_0 := \frac{1.99 - 2(\zeta_1+\zeta_2) - \sqrt{8/\pi}\alpha_h^{-1}}{4(1+\delta)^2 \cdot (1.02 + 2/\alpha_h)}$ and $\lambda + \mu \cdot 16(1+\delta)^2 \cdot (1.02 + 2/\alpha_h) < 4\left\{1.99 - 2(\zeta_1 + \zeta_2) - \sqrt{8/\pi}\alpha_h^{-1} - \epsilon\right\}$.

D. Geometric Convergence with Outliers (Proof of Theorem 1)

We consider the model (1) with outliers, i.e., $y_i = |\langle \boldsymbol{a}_i, \boldsymbol{x} \rangle|^2 + \eta_i$ for $i = 1, \cdots, m$. It suffices to show that $\nabla \ell_{tr}(z)$ satisfies the RC. The critical step is to lower and upper bound the sample median of the corrupted measurements. Lemma 3 yields

$$\theta_{\frac{1}{2}-s}(\{|(\boldsymbol{a}_i^T \boldsymbol{x})^2 - (\boldsymbol{a}_i^T \boldsymbol{z})^2|\}) \le \theta_{\frac{1}{2}}(\{|y_i - (\boldsymbol{a}_i^T \boldsymbol{z})^2|\}) \le \theta_{\frac{1}{2}+s}(\{|(\boldsymbol{a}_i^T \boldsymbol{x})^2 - (\boldsymbol{a}_i^T \boldsymbol{z})^2|\}. \tag{63}$$

For the instance of $s = 0.01$, by (37) in Proposition 2, we have with probability at least $1 - 2\exp(-\Omega(m)\epsilon^2)$,

$$(0.63 - \epsilon)\|\boldsymbol{z}\|\|\boldsymbol{h}\| \le \theta_{\frac{1}{2}}(\{|y_i - (\boldsymbol{a}_i^T\boldsymbol{z})^2|\}) \le (0.95 + \epsilon)\|\boldsymbol{z}\|\|\boldsymbol{h}\|. \tag{64}$$

To differentiate from $\mathcal{E}_2^i$, we define $\tilde{\mathcal{E}}_2^i := \left\{ \left|(\boldsymbol{a}_i^T\boldsymbol{x})^2 - (\boldsymbol{a}_i^T\boldsymbol{z})^2\right| \le \alpha_h \mathsf{med}\left\{\left|y_i - (\boldsymbol{a}_i^T\boldsymbol{z})^2\right|\right\} \frac{|\boldsymbol{a}_i^T\boldsymbol{z}|}{\|\boldsymbol{z}\|} \right\}$. We then have

$$\begin{aligned} -\nabla \ell_{tr}(\boldsymbol{z}) &= 2\sum_{i=1}^m \frac{(\boldsymbol{a}_i^T\boldsymbol{z})^2 - y_i}{\boldsymbol{a}_i^T\boldsymbol{z}} \boldsymbol{a}_i \mathbb{1}_{\mathcal{E}_1^i \cap \mathcal{E}_2^i} \\ &= \underbrace{2\sum_{i=1}^m \frac{(\boldsymbol{a}_i^T\boldsymbol{z})^2 - (\boldsymbol{a}_i^T\boldsymbol{x})^2}{\boldsymbol{a}_i^T\boldsymbol{z}} \boldsymbol{a}_i \mathbb{1}_{\mathcal{E}_1^i \cap \tilde{\mathcal{E}}_2^i}}_{\nabla^{clean}\ell_{tr}(\boldsymbol{z})} + \underbrace{2\sum_{i\in S}\left(\frac{(\boldsymbol{a}_i^T\boldsymbol{z})^2 - y_i}{\boldsymbol{a}_i^T\boldsymbol{z}} \mathbb{1}_{\mathcal{E}_1^i \cap \mathcal{E}_2^i} - \frac{(\boldsymbol{a}_i^T\boldsymbol{z})^2 - (\boldsymbol{a}_i^T\boldsymbol{x})^2}{\boldsymbol{a}_i^T\boldsymbol{z}} \mathbb{1}_{\mathcal{E}_1^i \cap \tilde{\mathcal{E}}_2^i} \right) \boldsymbol{a}_i}_{\nabla^{extra}\ell_{tr}(\boldsymbol{z})}. \end{aligned}$$

Choosing ϵ small enough, the inclusion property (i.e. $\mathcal{E}_3^i \subseteq \tilde{\mathcal{E}}_2^i \subseteq \mathcal{E}_4^i$) holds, and all the proof arguments for Proposition 3 and 4 are also valid to $\nabla^{clean}\ell_{tr}(\boldsymbol{z})$. Thus, one has

$$\frac{1}{m}\langle \nabla^{clean}\ell_{tr}(\boldsymbol{z}), \boldsymbol{h}\rangle \ge 2\left\{1.99 - 2(\zeta_1 + \zeta_2) - \sqrt{8/\pi}\alpha_h^{-1} - \epsilon\right\}\|\boldsymbol{h}\|^2, \tag{65}$$

$$\frac{1}{m}\left\|\nabla^{clean}\ell_{tr}(\boldsymbol{z})\right\| \le (1+\delta)\cdot 4\sqrt{1.02 + 2/\alpha_h}\|\boldsymbol{h}\|. \tag{66}$$

We next bound the contribution of $\nabla^{extra}\ell_{tr}(\boldsymbol{z})$. Introduce $\boldsymbol{q} = [q_1, \ldots, q_m]^T$, where

$$q_i := \left(\frac{(\boldsymbol{a}_i^T\boldsymbol{z})^2 - y_i}{\boldsymbol{a}_i^T\boldsymbol{z}} \mathbb{1}_{\mathcal{E}_1^i \cap \mathcal{E}_2^i} - \frac{(\boldsymbol{a}_i^T\boldsymbol{z})^2 - (\boldsymbol{a}_i^T\boldsymbol{x})^2}{\boldsymbol{a}_i^T\boldsymbol{z}} \mathbb{1}_{\mathcal{E}_1^i \cap \tilde{\mathcal{E}}_2^i} \right) \mathbb{1}_{\{i\in S\}}, \tag{67}$$

and then $|q_i| \le 2\alpha_h\|\boldsymbol{h}\|$. Thus $\|\boldsymbol{q}\| \le \sqrt{sm}\cdot 2\alpha_h\|\boldsymbol{h}\|$, and

$$\left\|\frac{1}{m}\nabla^{extra}\ell_{tr}(\boldsymbol{z})\right\| = \frac{1}{m}\left\|\boldsymbol{A}^T\boldsymbol{q}\right\| \le 2(1+\delta)\sqrt{s}\alpha_h\|\boldsymbol{h}\|, \tag{68}$$

$$\left|\left\langle \frac{1}{m}\nabla^{extra}\ell_{tr}(\boldsymbol{z}), \boldsymbol{h}\right\rangle\right| \le \|\boldsymbol{h}\|\cdot\left\|\frac{1}{m}\nabla^{extra}\ell_{tr}(\boldsymbol{z})\right\| \le 2(1+\delta)\sqrt{s}\alpha_h\|\boldsymbol{h}\|^2, \tag{69}$$

where $\boldsymbol{A} = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m]^T$. Then, we have

$$-\left\langle \frac{1}{m}\nabla\ell_{tr}(\boldsymbol{z}), \boldsymbol{h}\right\rangle \ge \left\langle \frac{1}{m}\nabla^{clean}\ell_{tr}(\boldsymbol{z}), \boldsymbol{h}\right\rangle - \left|\left\langle \frac{1}{m}\nabla^{extra}\ell_{tr}(\boldsymbol{z}), \boldsymbol{h}\right\rangle\right| \tag{70}$$

$$\ge 2\left(1.99 - 2(\zeta_1+\zeta_2) - \sqrt{8/\pi}\alpha_h^{-1} - \epsilon - (1+\delta)\sqrt{s}\alpha_h\right)\|\boldsymbol{h}\|^2, \tag{71}$$

and

$$\left\|\frac{1}{m}\nabla\ell_{tr}(\boldsymbol{z})\right\| \le \left\|\frac{1}{m}\nabla^{clean}\ell_{tr}(\boldsymbol{z})\right\| + \left\|\frac{1}{m}\nabla^{extra}\ell_{tr}(\boldsymbol{z})\right\| \tag{72}$$

$$\le (1+\delta)\left(4\sqrt{1.02 + 2/\alpha_h} + 2\sqrt{s}\alpha_h\right)\|\boldsymbol{h}\|. \tag{73}$$

The RC is guaranteed if μ, λ, ϵ are chosen properly and s is sufficiently small.

E. Geometric Convergence with Outliers and Bounded Noise (Proof of Theorem 2)

We consider the model (2) with outliers and bounded noise, i.e., $y_i = |\langle \boldsymbol{a}_i, \boldsymbol{x}\rangle|^2 + w_i + \eta_i$ for $i = 1, \cdots, m$. We omit the initialization analysis as it is similar to Appendix B. We split our analysis of the gradient loop into two regimes.

• **Regime 1**: $c_4\|\boldsymbol{z}\| \geq \|\boldsymbol{h}\| \geq c_3 \frac{\|\boldsymbol{w}\|_\infty}{\|\boldsymbol{z}\|}$. In this regime, error contraction by each gradient step is given by

$$\text{dist}\left(\boldsymbol{z} + \frac{\mu}{m}\nabla\ell_{tr}(\boldsymbol{z}), \boldsymbol{x}\right) \leq (1-\rho)\text{dist}(\boldsymbol{z}, \boldsymbol{x}). \tag{74}$$

It suffices to justify that $\nabla\ell_{tr}(\boldsymbol{z})$ satisfies the RC. Denote $\tilde{y}_i := (\boldsymbol{a}_i^T\boldsymbol{x})^2 + w_i$. Then by Lemma 3, we have

$$\theta_{\frac{1}{2}-s}\left\{\left|\tilde{y}_i - (\boldsymbol{a}_i^T\boldsymbol{z})^2\right|\right\} \leq \text{med}\left\{\left|y_i - (\boldsymbol{a}_i^T\boldsymbol{z})^2\right|\right\} \leq \theta_{\frac{1}{2}+s}\left\{\left|\tilde{y}_i - (\boldsymbol{a}_i^T\boldsymbol{z})^2\right|\right\}. \tag{75}$$

Moreover, by Lemma 2 we have

$$\left|\theta_{\frac{1}{2}+s}\left\{\left|\tilde{y}_i - (\boldsymbol{a}_i^T\boldsymbol{z})^2\right|\right\} - \theta_{\frac{1}{2}+s}\left\{\left|(\boldsymbol{a}_i^T\boldsymbol{x})^2 - (\boldsymbol{a}_i^T\boldsymbol{z})^2\right|\right\}\right| \leq \|\boldsymbol{w}\|_\infty, \tag{76}$$

$$\left|\theta_{\frac{1}{2}-s}\left\{\left|\tilde{y}_i - (\boldsymbol{a}_i^T\boldsymbol{z})^2\right|\right\} - \theta_{\frac{1}{2}-s}\left\{\left|(\boldsymbol{a}_i^T\boldsymbol{x})^2 - (\boldsymbol{a}_i^T\boldsymbol{z})^2\right|\right\}\right| \leq \|\boldsymbol{w}\|_\infty. \tag{77}$$

Assume that $s = 0.01$ and apply Proposition 2. Moreover, if c_3 is sufficiently large (i.e., $c_3 > 100$) and ϵ is small enough (i.e., $\epsilon < 0.02$), then we have

$$0.6\|\boldsymbol{x}-\boldsymbol{z}\|\|\boldsymbol{z}\| \leq \text{med}\left\{\left|y_i - (\boldsymbol{a}_i^T\boldsymbol{z})^2\right|\right\} \leq 1\|\boldsymbol{x}-\boldsymbol{z}\|\|\boldsymbol{z}\|. \tag{78}$$

Furthermore, recall $\tilde{\mathcal{E}}_2^i := \left\{\left|(\boldsymbol{a}_i^T\boldsymbol{x})^2 - (\boldsymbol{a}_i^T\boldsymbol{z})^2\right| \leq \alpha_h \text{med}\left\{\left|(\boldsymbol{a}_i^T\boldsymbol{z})^2 - y_i\right|\right\}\frac{|\boldsymbol{a}_i^T\boldsymbol{z}|}{\|\boldsymbol{z}\|}\right\}$, then

$$\begin{aligned}
-\nabla\ell_{tr}(\boldsymbol{z}) &= 2\sum_{i=1}^m \frac{(\boldsymbol{a}_i^T\boldsymbol{z})^2 - y_i}{\boldsymbol{a}_i^T\boldsymbol{z}}\boldsymbol{a}_i\mathbb{1}_{\mathcal{E}_1^i\cap\mathcal{E}_2^i} \\
&= \underbrace{2\left(\sum_{i\notin S}\frac{(\boldsymbol{a}_i^T\boldsymbol{z})^2 - (\boldsymbol{a}_i^T\boldsymbol{x})^2}{\boldsymbol{a}_i^T\boldsymbol{z}}\boldsymbol{a}_i\mathbb{1}_{\mathcal{E}_1^i\cap\mathcal{E}_2^i} + \sum_{i\in S}\frac{(\boldsymbol{a}_i^T\boldsymbol{z})^2 - (\boldsymbol{a}_i^T\boldsymbol{x})^2}{\boldsymbol{a}_i^T\boldsymbol{z}}\boldsymbol{a}_i\mathbb{1}_{\mathcal{E}_1^i\cap\tilde{\mathcal{E}}_2^i}\right)}_{\nabla^{clean}\ell_{tr}(\boldsymbol{z})} \\
&\quad \underbrace{-2\sum_{i\notin S}\frac{\boldsymbol{w}_i}{\boldsymbol{a}_i^T\boldsymbol{z}}\boldsymbol{a}_i\mathbb{1}_{\mathcal{E}_1^i\cap\mathcal{E}_2^i}}_{\nabla^{noise}\ell_{tr}(\boldsymbol{z})} + \underbrace{2\sum_{i\in S}\left(\frac{(\boldsymbol{a}_i^T\boldsymbol{z})^2 - y_i}{\boldsymbol{a}_i^T\boldsymbol{z}}\mathbb{1}_{\mathcal{E}_1^i\cap\mathcal{E}_2^i} - \frac{(\boldsymbol{a}_i^T\boldsymbol{z})^2 - (\boldsymbol{a}_i^T\boldsymbol{x})^2}{\boldsymbol{a}_i^T\boldsymbol{z}}\mathbb{1}_{\mathcal{E}_1^i\cap\tilde{\mathcal{E}}_2^i}\right)\boldsymbol{a}_i}_{\nabla^{extra}\ell_{tr}(\boldsymbol{z})}.
\end{aligned}$$

For $i \notin S$, the inclusion property (i.e. $\mathcal{E}_3^i \subseteq \mathcal{E}_2^i \subseteq \mathcal{E}_4^i$) holds because

$$\left|y_i - (\boldsymbol{a}_i^T\boldsymbol{z})^2\right| \in \left|(\boldsymbol{a}_i^T\boldsymbol{x})^2 - (\boldsymbol{a}_i^T\boldsymbol{z})^2\right| \pm |w_i|$$

and $|w_i| \leq \frac{1}{c_3}\|\boldsymbol{h}\|\|\boldsymbol{z}\|$ for some sufficient large c_3. For $i \in S$, the inclusion $\mathcal{E}_3^i \subseteq \tilde{\mathcal{E}}_2^i \subseteq \mathcal{E}_4^i$ holds because of (78). All the proof arguments for Proposition 3 and 4 are also valid for $\nabla^{clean}\ell_{tr}(\boldsymbol{z})$, and thus we have

$$\frac{1}{m}\langle\nabla^{clean}\ell_{tr}(\boldsymbol{z}), \boldsymbol{h}\rangle \geq 2\left\{1.99 - 2(\zeta_1+\zeta_2) - \sqrt{8/\pi}\alpha_h^{-1} - \epsilon\right\}\|\boldsymbol{h}\|^2, \tag{79}$$

$$\frac{1}{m}\left\|\nabla^{clean}\ell_{tr}(\boldsymbol{z})\right\| \leq (1+\delta)\cdot 4\sqrt{1.02 + 2/\alpha_h}\|\boldsymbol{h}\|. \tag{80}$$

Next, we turn to control the contribution of the noise. Let $\tilde{\boldsymbol{w}}_i = \frac{2\boldsymbol{w}_i}{\boldsymbol{a}_i^T\boldsymbol{z}}\mathbb{1}_{\mathcal{E}_1^i\cap\mathcal{E}_2^i}$, then we have

$$\frac{1}{m}\|\nabla^{noise}\ell_{tr}(\boldsymbol{z})\| = \left\|\frac{1}{m}\boldsymbol{A}^T\tilde{\boldsymbol{w}}\right\| \leq \left\|\frac{1}{\sqrt{m}}\boldsymbol{A}^T\right\|\left\|\frac{\tilde{\boldsymbol{w}}}{\sqrt{m}}\right\| \leq (1+\delta)\|\tilde{\boldsymbol{w}}\|_\infty \leq (1+\delta)\frac{2\|\boldsymbol{w}\|_\infty}{\alpha_l\|\boldsymbol{z}\|}, \tag{81}$$

when m/n is sufficiently large. Given the regime condition $\|\boldsymbol{h}\| \geq c_3\frac{\|\boldsymbol{w}\|_\infty}{\|\boldsymbol{z}\|}$, we further have

$$\frac{1}{m}\|\nabla^{noise}\ell_{tr}(\boldsymbol{z})\| \leq \frac{2(1+\delta)}{c_3\alpha_l}\|\boldsymbol{h}\|, \tag{82}$$

$$\frac{1}{m}\left|\langle\nabla^{noise}\ell_{tr}(\boldsymbol{z}), \boldsymbol{h}\rangle\right| \leq \frac{1}{m}\left\|\nabla^{noise}\ell_{tr}(\boldsymbol{z})\right\|\cdot\|\boldsymbol{h}\| \leq \frac{2(1+\delta)}{c_3\alpha_l}\|\boldsymbol{h}\|^2. \tag{83}$$

We next bound the contribution of $\nabla^{extra}\ell_{tr}(\boldsymbol{z})$. Introduce $\boldsymbol{q} = [q_1, \ldots, q_m]^T$, where

$$q_i := 2\left(\frac{(\boldsymbol{a}_i^T\boldsymbol{z})^2 - y_i}{\boldsymbol{a}_i^T\boldsymbol{z}}\mathbb{1}_{\mathcal{E}_1^i \cap \mathcal{E}_2^i} - \frac{(\boldsymbol{a}_i^T\boldsymbol{z})^2 - (\boldsymbol{a}_i^T\boldsymbol{x})^2}{\boldsymbol{a}_i^T\boldsymbol{z}}\mathbb{1}_{\mathcal{E}_1^i \cap \tilde{\mathcal{E}}_2^i}\right)\mathbb{1}_{\{i\in S\}}. \tag{84}$$

Then $|q_i| \le 2\alpha_h\|\boldsymbol{h}\|$, and $\|\boldsymbol{q}\| \le \sqrt{sm}\cdot 2\alpha_h\|\boldsymbol{h}\|$. We thus have

$$\left\|\frac{1}{m}\nabla^{extra}\ell_{tr}(\boldsymbol{z})\right\| = \frac{1}{m}\left\|\boldsymbol{A}^T\boldsymbol{q}\right\| \le 2(1+\delta)\sqrt{s}\alpha_h\|\boldsymbol{h}\|, \tag{85}$$

$$\left|\left\langle\frac{1}{m}\nabla^{extra}\ell_{tr}(\boldsymbol{z}), \boldsymbol{h}\right\rangle\right| \le \|\boldsymbol{h}\|\cdot\left\|\frac{1}{m}\nabla^{extra}\ell_{tr}(\boldsymbol{z})\right\| \le 2(1+\delta)\sqrt{s}\alpha_h\|\boldsymbol{h}\|^2. \tag{86}$$

Putting these together, one has

$$\begin{aligned}-\frac{1}{m}\left\langle\nabla\ell_{tr}(\boldsymbol{z}), \boldsymbol{h}\right\rangle &\ge \frac{1}{m}\left\langle\nabla^{clean}\ell_{tr}(\boldsymbol{z}), \boldsymbol{h}\right\rangle - \frac{1}{m}\left|\left\langle\nabla^{noise}\ell_{tr}(\boldsymbol{z}), \boldsymbol{h}\right\rangle\right| - \frac{1}{m}\left|\left\langle\nabla^{extra}\ell_{tr}(\boldsymbol{z}), \boldsymbol{h}\right\rangle\right| \\ &\ge 2\left(1.99 - 2(\zeta_1+\zeta_2) - \sqrt{8/\pi}\alpha_h^{-1} - \epsilon - (1+\delta)(1/(c_3\alpha_z^l) + \sqrt{s}\alpha_h)\right)\|\boldsymbol{h}\|^2,\end{aligned} \tag{87}$$

and

$$\begin{aligned}\frac{1}{m}\left\|\nabla\ell_{tr}(\boldsymbol{z})\right\| &\le \frac{1}{m}\left\|\nabla^{clean}\ell_{tr}(\boldsymbol{z})\right\| + \frac{1}{m}\left\|\nabla^{noise}\ell_{tr}(\boldsymbol{z})\right\| + \frac{1}{m}\left\|\nabla^{extra}\ell_{tr}(\boldsymbol{z})\right\| \\ &\le 2(1+\delta)\left(2\sqrt{1.02 + 2/\alpha_h} + 1/(c_3\alpha_z^l) + \sqrt{s}\alpha_h\right)\|\boldsymbol{h}\|.\end{aligned} \tag{88}$$

The RC is guaranteed if μ, λ, ϵ are chosen properly, c_3 is sufficiently large and s is sufficiently small.

• **Regime 2**: Once the iterate enters this regime with $\|\boldsymbol{h}\| \le \frac{c_3\|\boldsymbol{w}\|_\infty}{\|\boldsymbol{z}\|}$, each gradient iterate may not reduce the estimation error. However, in this regime each move size $\frac{\mu}{m}\nabla\ell_{tr}(\boldsymbol{z})$ is at most $\mathcal{O}(\|\boldsymbol{w}\|_\infty/\|\boldsymbol{z}\|)$. Then the estimation error cannot increase by more than $\frac{\|\boldsymbol{w}\|_\infty}{\|\boldsymbol{z}\|}$ with a constant factor. Thus one has

$$\text{dist}\left(\boldsymbol{z} + \frac{\mu}{m}\nabla\ell_{tr}(\boldsymbol{z}), \boldsymbol{x}\right) \le c_5\frac{\|\boldsymbol{w}\|_\infty}{\|\boldsymbol{x}\|} \tag{89}$$

for some constant c_5. As long as $\|\boldsymbol{w}\|_\infty/\|\boldsymbol{x}\|^2$ is sufficiently small, it is guaranteed that $c_5\frac{\|\boldsymbol{w}\|_\infty}{\|\boldsymbol{x}\|} \le c_4\|\boldsymbol{x}\|$. If the iterate jumps out of *Regime 2*, it falls into *Regime 1*.

F. Proof of Proposition 3

The proof adapts the proof of Proposition 2 in (Chen & Candès, 2015). We outline the main steps for completeness. Observe that for the noise-free case, $y_i = (\boldsymbol{a}_i^T\boldsymbol{x})^2$. We obtain

$$\begin{aligned}-\frac{1}{2m}\nabla\ell_{tr}(\boldsymbol{z}) &= \frac{1}{m}\sum_{i=1}^m\frac{(\boldsymbol{a}_i^T\boldsymbol{z})^2 - (\boldsymbol{a}_i^T\boldsymbol{x})^2}{\boldsymbol{a}_i^T\boldsymbol{z}}\boldsymbol{a}_i\mathbb{1}_{\mathcal{E}_1^i\cap\mathcal{E}_2^i} \\ &= \frac{1}{m}\sum_{i=1}^m 2(\boldsymbol{a}_i^T\boldsymbol{h})\boldsymbol{a}_i\mathbb{1}_{\mathcal{E}_1^i\cap\mathcal{E}_2^i} - \frac{1}{m}\sum_{i=1}^m\frac{(\boldsymbol{a}_i^T\boldsymbol{h})^2}{\boldsymbol{a}_i^T\boldsymbol{z}}\boldsymbol{a}_i\mathbb{1}_{\mathcal{E}_1^i\cap\mathcal{E}_2^i}.\end{aligned} \tag{90}$$

One expects the contribution of the second term in (90) to be small as $\|\boldsymbol{h}\|/\|\boldsymbol{z}\|$ decreases.

Specifically, following the two inclusion properties (54) and (58), we have

$$\mathcal{D}_{\gamma_3}^{i,1}\cap\mathcal{E}_{1,\gamma_3}^i \subseteq \mathcal{E}_3^i\cap\mathcal{E}_1^i \subseteq \mathcal{E}_2^i\cap\mathcal{E}_1^i \subseteq \mathcal{E}_4^i\cap\mathcal{E}_1^i \subseteq (\mathcal{D}_{\gamma_4}^{i,1}\cup\mathcal{D}_{\gamma_4}^{i,2})\cap\mathcal{E}_1^i \tag{91}$$

where the parameters γ_3, γ_4 are given by

$$\gamma_3 := 0.248\alpha_h, \quad \text{and} \quad \gamma_4 := \alpha_h. \tag{92}$$

Continuing with the identity (90), we have a lower bound

$$-\left\langle \frac{1}{2m}\nabla \ell_{tr}(\boldsymbol{z}), \boldsymbol{h} \right\rangle \geq \frac{2}{m}\sum_{i=1}^{m}(\boldsymbol{a}_i^T\boldsymbol{h})^2 \mathbb{1}_{\mathcal{E}_1^i \cap \mathcal{D}_{\gamma_3}^{i,1}} - \frac{1}{m}\sum_{i=1}^{m}\frac{|\boldsymbol{a}_i^T\boldsymbol{h}|^3}{|\boldsymbol{a}_i^T\boldsymbol{z}|}\mathbb{1}_{\mathcal{D}_{\gamma_4}^{i,1}\cap \mathcal{E}_1^i} - \frac{1}{m}\sum_{i=1}^{m}\frac{|\boldsymbol{a}_i^T\boldsymbol{h}|^3}{|\boldsymbol{a}_i^T\boldsymbol{z}|}\mathbb{1}_{\mathcal{D}_{\gamma_4}^{i,2}\cap \mathcal{E}_1^i}. \tag{93}$$

The three terms in (93) can be bounded following Lemmas 4, 5, and 6 in (Chen & Candès, 2015), which concludes the proof.